A Systematic Test of the Independence Axiom Near Certainty*

Ritesh Jain† Kirby Nielsen‡

July 7, 2020

Abstract

A large literature has documented violations of expected utility consistent with a preference for certainty (the "certainty effect"). We design a laboratory experiment to investigate the role of the certainty effect in explaining violations of the independence axiom. We use lotteries spanning over the entire probability simplex to detect violations systematically. We find that violations of independence consistent with the *reverse* certainty effect are much more common than violations consistent with the certainty effect. Results hold as we test robustness along two dimensions: varying the mixing lottery and moving slightly away from certainty.

Keywords: independence axiom; expected utility theory; certainty effect; Allais Paradox

JEL classification: C79, D82

*We want to thank Yaron Azrieli, Doug Bernheim, Anujit Chakraborty, Paul J. Healy, Collin Raymond, John Rehbeck, Colin Sullivan, and the participants of the experimental reading group at the Ohio State University for providing helpful feedback on this project. All errors remain our own. Research support was provided by the Decision Science Collaborative of the Ohio State University. This study was approved by the Institutional Review Boards at Academia Sinica and Stanford University.

†Institute of Economics, Academia Sinica; *ritesh@econ.sinica.edu.tw*

‡Department of Economics, Stanford University; *kirbyn@stanford.edu*

"Consequently, I viewed the principle of independence as incompatible with the preference for security in the neighborhood of certainty shown by every subject... This led me to devise some counter-examples. One of them, formulated in 1952, has become famous as the 'Allais Paradox.' Today, it is as widespread as its real meaning is generally misunderstood." — Maurice Allais

I. INTRODUCTION

Experimental evidence has shown that individuals reliably violate the independence axiom, the central tenet of expected utility theory.[1] In 1952, Maurice Allais proposed one of the earliest, and still to-date most famous, counter-examples, now known as the "Allais Paradox." For concreteness, consider the common ratio version of the Allais Paradox introduced by Kahneman and Tversky (1979). Ask a decision maker the following two binary questions:

	Option A:		Option B:
Q1:	100% chance of \$3000	vs.	80% chance of \$4000 20% chance of \$0
Q2:	25% chance of \$3000 75% chance of \$0	vs.	20% chance of \$4000 80% chance of \$0

Many individuals choose Option A in the first decision and choose Option B in the second.[2] This pattern of choices violates the independence axiom.[3]

Allais attributed these violations to a *preference for security*, quoted above, now referred to as the "certainty effect" (Kahneman and Tversky, 1979). Kahneman and Tversky describe the certainty effect as the phenomenon in which "people overweight outcomes that are considered certain, relative to outcomes which are merely probable" (Kahneman and Tversky, 1979, p. 265). The intuition in the Allais Paradox is that the preference of Option A over Option B in Q1 is driven, in part, by the fact that Option A offers a sure payoff. When both options are risky, as in Q2, neither offers

[1]Independence states that for any three lotteries p, q, and r, and any number λ in [0,1], if p is preferred to q, then $\lambda p+(1-\lambda)r$ is preferred to $\lambda q+(1-\lambda)r$. That is, mixing both lotteries p and q with a common lottery r, and in common proportions, should not change the relative preference between p and q.

[2]Kahneman and Tversky (1979) find that 80% of individuals preferred Option A in Question 1, while 65% preferred Option B in Question 2. Over half of their subjects violated independence.

[3]To see this, let $\lambda = 0.25$ and r be 100% chance of \$0.

the appeal of certainty, so preferences can reverse. Allais's original intuition, shared by many and confirmed by experimental evidence, has led to large theoretical and experimental literatures in search of a descriptive non-expected utility model. We review these papers in Section II. The certainty effect also has been invoked to explain behaviors outside the domain of simple lotteries, such as present bias (Halevy, 2008) and aversion to gradual pieces of information (Dillenberger, 2010).

Given the prominence of the Allais Paradox and robust evidence of the certainty effect, a persistent thread in the literature is that the certainty effect drives violations of independence. For example, Schmidt (1998) says that "the bulk of observed violations of the independence axiom is due to the certainty effect." Under this view, the claim is that most violations of independence occur when an individual prefers a sure amount to a risky lottery. A few papers have found significant violations of independence when the risky lottery is preferred to the sure amount before mixing—the opposite pattern, or "reverse certainty effect"—but there has been no systematic test of the independence axiom and the certainty effect.

The goal of our paper is to provide such a test. We fix a probability simplex, which, in our experiment, is the set of possible lotteries over {\$10, \$20, \$30}. We pick forty-five lotteries uniformly across this simplex. Subjects face binary choices between \$20 for sure and a randomly-selected subset of these uniformly-distributed risky lotteries. Given the wide range of lotteries we sample, subjects will prefer \$20 to the risky lottery in some questions, while in other questions they will prefer the risky lottery to \$20. We then mix the alternatives according to three different probabilities to see if preferences reverse, constituting a violation of independence. This allows us to detect and compare independence violations when certainty is preferred to risk (i.e. when individuals prefer \$20 for sure to the risky lottery, but then reverse preferences once both are mixed), consistent with the certainty effect, to those when risk is preferred to certainty (i.e. when individuals prefer the risky lottery to \$20 for sure, but then reverse preferences when both are mixed), consistent with the reverse certainty effect.

We find that *reverse* certainty effect violations are far more common than certainty effect violations. Conditional on preferring certainty to risk, individuals violate independence 15% of the time. In stark contrast, individuals violate independence almost 40% of the time conditional on preferring risk to certainty. Results hold on an individual level, as well. The modal subject never violates independence in questions where they prefer certainty to a risky lottery, while the modal subject violates independence in one-third of questions where they prefer the risky lottery over certainty. Our design also allows us to look at a more rigorous test of the independence axiom, comparing choices across all four mixing probabilities. This analysis confirms our main results, with most violations coming from instances where individuals choose the risky lottery over certainty and then switch to choosing the safer lottery as the alternatives are mixed.

We test the robustness of these results along two dimensions. First, we move slightly away from certainty by comparing with a lottery which gives \$20 with 90% chance (otherwise a 5% chance of \$30 and a 5% chance of \$10) rather than \$20 with certainty. Second, we vary the "mixing lottery" from one in the spirit of the Allais Paradox to one less commonly studied. Overall, our results are robust to these perturbations.

Our results contribute to a large experimental literature testing the independence axiom. Ours is the first large-scale systematic test around certainty, giving general evidence on the frequency and location of independence violations. We believe our results will be particularly useful to incorporate into theoretical models of choice under risk. Recent theories seek to characterize and axiomatize the certainty effect in building descriptive models of choice. Our results suggest that these theories may miss an important pattern of behavior. In our data, we could explain significantly more choices by modeling the exact opposite preference.

II. Literature Review

Since Allais's objections to the descriptive validity of the Expected Utility Theory (EUT), an enormous amount of theoretical effort has been devoted to developing alternatives to EUT. Hand-in-hand with theoretical advancements is an experimental program aimed at testing these theories. The experimental literature is vast, so we cannot summarize every paper here. Camerer (1995) and Starmer (2000) review the older literature, and we refer the interested reader to those surveys.

In a recent meta-analysis, Blavatskyy et al. (2015) analyze results from 39 common consequence Allais Paradox experiments, which is one specific pattern of independence axiom violations. In the original formulation of the common consequence effect, we ask a decision maker to choose in the following binary decisions:

	Option A:		Option B:
Q1:	100% chance of \$100 million	vs.	10% chance of \$500 million 89% chance of \$100 million 1% chance of \$0
Q2:	11% chance of \$100 million 89% chance of \$0	vs.	10% chance of \$500 million 90% chance of \$0

Many decision makers choose Option A in Q1 and Option B in Q2, which constitutes a violation of independence. Blavatskyy et al. (2015) analyze results from a number of such questions that vary on many dimensions including size of lottery stakes, real vs. hypothetical prizes, mixing probabilities, etc. Most papers ask only a single pair of questions involving certainty.

Blavatskyy et al. look to see whether violations are more consistent with the certainty effect or the reverse certainty effect.[4] In 22 out of 39 instances, the certainty effect is more common. In six instances, the reverse certainty effect is more common (Conlisk, 1989; Starmer, 1992; Humphrey and Verschoor, 2004; Blavatskyy, 2013). There is no statistically significant difference between the two types of violations in the remaining 11 instances.

As Blavatskyy et al. (2015) discuss, the Allais Paradox, therefore, seems less prevalent than one might expect given its prominence in the literature. Since the papers reviewed in the meta-analysis vary on many dimensions, there is no systematic way to compare the evidence. While there has been sparse evidence of the reverse certainty effect (in only five incentivized questions), no paper has looked at the whole simplex in a structured way. The main contribution of our paper is to ask subjects a large number of questions across the simplex involving certainty, or near certainty, to study where violations of independence emerge.

The prominence of the certainty effect in certain strands of the literature has led to theoretical work attempting to capture the empirical evidence. These alternatives to EUT typically weaken the independence axiom to accommodate the Allais Paradox. Some of the popular alternative theories are disappointment aversion (Gul, 1991), cumulative prospect theory (Tversky and Kahneman, 1992), rank dependent utility theory (Quiggin, 1982), weighted expected utility theory (Hong and Waller, 1986), implicit expected utility theory (Dekel, 1986), and cautious expected utility theory

[4] In their Table 1, the certainty effect is labeled "SR" and the reverse certainty effect is labeled "RS."

(Cerreia-Vioglio et al., 2015).[5]

Each non-EU theory listed above can accommodate the certainty effect, but we highlight here Cautious Expected Utility (Cerreia-Vioglio et al., 2015) because it was designed exactly to characterize certainty-effect preferences. Cerreia-Vioglio et al. weaken independence by requiring it to hold only when risk is already preferred to certainty, allowing for independence violations when certainty is preferred to risk. Formally, they replace independence with an axiom, Negative Certainty Independence (NCI), first introduced in Dillenberger (2010). NCI states that for all lotteries $p, q \in \Delta(X)$, prizes $x \in X$, degenerate lotteries δ_x, and probabilities $\lambda \in [0,1]$,

$$p \succeq \delta_x \Rightarrow \lambda p + (1-\lambda)q \succeq \lambda \delta_x + (1-\lambda)q.$$

This requires that independence holds when a lottery, p, is preferred to $\$x$ for sure, but does not require independence to hold in the opposite case where certainty is preferred to risk. As such, this theory exactly characterizes the preference for certainty underlying the Allais paradox and other commonly-observed patterns of behavior.

However, the authors point out that "no comprehensive tests of NCI have been conducted thus far" (Cerreia-Vioglio et al. (2015), p. 713). We see our paper as a natural step in this dialogue between theory and experiments. Our results suggest that the certainty effect is not always the main obstacle for the independence axiom—in our data, the *reverse* certainty effect is the main obstacle for independence, emphasizing the importance of a systematic test of independence to inform behaviorally-descriptive theories. Indeed, in a more recent paper, Cerreia-Vioglio et al. (2020) characterize preferences with the opposite axiom, Positive Certainty Independence (PCI), which requires independence hold instead when a sure payment is preferred to a risky lottery.

Finally, a few recent papers in various domains suggest that the certainty effect requires true certainty (probability one), and differs predictably from "near certainty" (Halevy, 2008; Andreoni and Harbaugh, 2010; Andreoni and Sprenger, 2010, 2011, 2012). We include tests of independence involving true certainty as well as tests near certainty. This allows us to see whether the patterns we find are robust to this perturbation, and allows us to test whether independence is violated more often in

[5] While many of these theories can accommodate the reverse certainty effect in the "opposite" way, they were designed to capture the certainty effect.

questions that offer certainty. Our main results hold both under certainty and near-certainty.

III. Theoretical Framework

We describe the theoretical framework in the context of our experimental design. All questions involve lotteries over US dollars. The set of possible prizes in our experiment is $X = \{10, 20, 30\}$. We represent the set of lotteries with prizes in X by $\Delta(X)$, with weak preferences $\succeq$ defined over $\Delta(X)$. We denote generic prizes in X by x, y, z, and denote generic lotteries in $\Delta(X)$ by p, q, r, s. The probability of receiving prize x under lottery p is denoted $p(x)$. We represent the three-outcome lottery, p, giving \$10 with probability $p(10)$, \$20 with probability $p(20)$, and \$30 with probability $p(30)$ by $(\$30, p(30); \$20, p(20); \$10, p(10))$. We represent the degenerate lottery giving $\$x$ for sure as δ_x.

The independence axiom states that for all $p, q, r \in \Delta(X)$ and for all $\lambda \in [0,1]$,

$$p \succeq q \Leftrightarrow \lambda p + (1-\lambda) r \succeq \lambda q + (1-\lambda) r.$$

We consider only "one-stage" lottery mixtures, rather than two-stage compound lotteries.[6] In our experiment, we will test the independence axiom by presenting subjects with binary choices over these one-stage lotteries.

There are two ways individuals can violate independence when one option is certain. The *certainty effect* (CE) captures the idea that individuals place disproportionate weight on an outcome when it is certain (Kahneman and Tversky, 1979). Individuals with a strong preference for certainty will be more likely to violate independence when certainty is preferred to a risky lottery before mixing. The intuition is that the preference of δ_x over p may be driven, in part, by the certainty appeal of receiving $\$x$ for sure. When these lotteries are mixed as in independence, $\lambda \delta_x + (1-\lambda) r$ does not carry the same certainty appeal, which might result in a preference for $\lambda p + (1-\lambda) r$ over $\lambda \delta_x + (1-\lambda) r$. When individuals violate independence in this way, we call it a "CE" violation.

The *reverse certainty effect* (RCE) is the exact opposite pattern. This refers to an individual who chooses p over δ_x and then chooses $\lambda \delta_x + (1-\lambda) r$ over $\lambda p + (1-\lambda) r$.

[6] In other words, we study *mixture independence*, rather than *compound independence*, as defined in Segal (1990).

We refer to this as an "RCE" violation. Our main research question is documenting the prevalence of independence violations and comparing the frequency of these two patterns of violations in a systematic way.

IV. Experimental Design

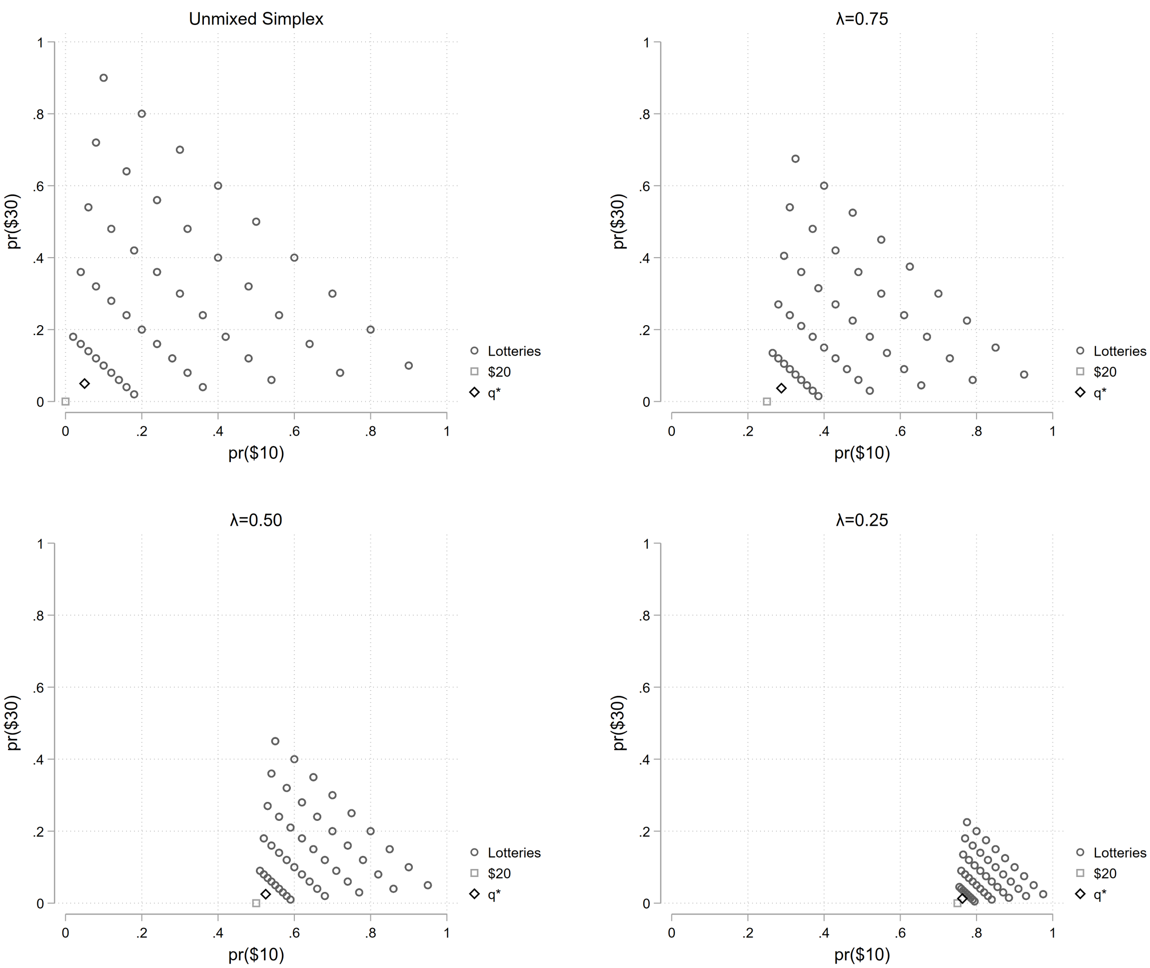


Figure I: Questions

We chose three payments—$10, $20, and $30—and all questions involve lotteries over these three payments.[7] In order to compare CE and RCE violations, we needed

[7]There is evidence that independence violations under certainty are more prevalent with large stakes than small stakes, reviewed in Cerreia-Vioglio et al. (2015). Therefore, we wanted to pick payments that were fairly high. Our payments averaged to around $20 per person, and sessions took only 30 minutes. Subjects knew this ahead of time. We felt this $40/hr average payment would be reasonably high stakes based on the literature.

to ask questions where a risky lottery is likely to be preferred to certainty, as well as questions where certainty is likely to be preferred to the risky lottery. To ensure this, we selected 45 points uniformly across the simplex. These 45 questions are denoted with circles in the top left graph of Figure I, and we refer to these as the "unmixed lotteries." We asked binary questions comparing these lotteries against a sure payment of \$20: a choice of p vs. δ_{20}.

To test independence, we mixed these lotteries with $r = (\$10, 1)$. We used three different mixing probabilities, $\lambda = \{0.25, 0.50, 0.75\}$. This results in 45 new binary choices for each value of λ: λp vs. $\lambda\delta_{20}$. The lotteries after mixing are shown in the remaining three panels of Figure I.[8]

We test the robustness of the certainty effect by moving δ_{20} slightly away from certainty, comparing these unmixed lotteries against $(\$30, 0.05; \$20, 0.90; \$10, 0.05)$, denoted by a diamond in Figure I. This lottery is "close" to a sure payment of \$20, but does not offer the same security. For simplicity, we'll call this lottery q^*, and we'll refer to these questions as "near-certain." Subjects face both certain and near-certain questions, as we explain below. When talking about alternatives in these binary comparisons, we refer to p and λp as the "risky lotteries" and refer to either δ_{20} and $\lambda\delta 20$, or q^* and λq^*, as the "safer lotteries." We reserve "certainty" only for δ_{20}.

In total, we have 360 possible questions–the 45 unmixed lotteries compared with \$20 in the *certain* condition (45 questions) and compared with q^* in the *near-certain* condition (45 questions). These 90 questions comprise the "unmixed" comparisons, and each is mixed by $\lambda = 0.75, 0.50, 0.25$ $(90 \times 4 = 360)$. Since it might be unreasonable for individuals to answer all 360 questions, each subject instead answered 68 binary questions from the set of 360 possible questions.[9] To perform the random selection, we created a bank of 90 questions—the 45 unmixed lotteries compared against \$20 and the same 45 unmixed lotteries compared against q^*. We randomly and independently selected 17 of these 90 questions for each subject. For those 17 questions, we asked subjects the unmixed question and all three $\lambda = \{0.25, 0.50, 0.75\}$ mixtures. This gives a total of $17 \times 4 = 68$ binary choices per subject.

This random selection process helps ensure that, on average for each subject, we

[8]Though we sometimes refer to lotteries "before" or "after" mixing for ease of exposition, there is no temporal component to the experiment. As we explain, questions were presented to subjects in random order.

[9]68 was calibrated based on duration of the experiment.

will have observations where the risky lottery is preferred to the safer lottery and vice versa, and we will also have observations for both certain and near-certain comparisons. It also allows us to test independence more rigorously than in single binary choices, as independence requires an individual to choose either the risky or safer option in all four λ comparisons. This design also rules out the possibility that independence violations result from indifference, which is a common critique of experiments that observe preference reversals (Blavatskyy, 2010). Given the number and diversity of questions we ask, systematic and persistent violations of independence cannot be explained through indifference.[10]

Finally, we conducted two between-subject treatments. The first, which we have explained above, mixes lotteries with the bottom right of the simplex, $r = (\$10, 1)$. This is closest in spirit to the original Allais Paradox where the lotteries were mixed with the lowest possible payoff. We refer to this as the "Allais Mix" treatment. To further test the robustness of independence violations, we ran a separate treatment that mixes lotteries instead with the midpoint of the simplex, $r = (30, \frac{1}{3}; 20, \frac{1}{3}; 10, \frac{1}{3})$, which we refer to as the "Middle Mix" treatment. Each subject participated in either the Allais Mix or Middle Mix treatment, but not both. We defer explanation of the Middle Mix treatment to Section V.

Procedures

We present results from 14 experimental sessions with a total of 265 subjects, 118 in the Allais Mix treatment and 147 in the Middle Mix treatment. Subjects were mainly undergraduates from Ohio State University, recruited using ORSEE (Greiner, 2004). The experiment was programmed using z-Tree (Fischbacher, 2007). Sessions lasted approximately 30 minutes and subject payments averaged \$20.

The experimenter read instructions out loud to all subjects. Instructions explained the binary choices and how the probabilities would translate into payoffs. Computer screens displayed the written probabilities and payoffs, as well as color-coded pie charts. Figure VIII in the Appendix shows a screenshot. All 68 questions were displayed in random order, randomized separately across subjects. In particular, it was not necessarily the case that subjects first saw the unmixed question, then the $\lambda = 0.75, 0.5, 0.25$ questions, and subjects were unaware that questions were related

[10] Given the structure of our lotteries, subjects could be exactly indifferent to \$20 on one "ray" from the origin, which is at most 5 questions.

in any way. Furthermore, each question was displayed on a separate screen, and we randomized the left-right screen position of the risky and safer lottery.

Subjects were paid after everyone in the session completed the experiment. We used physical randomization devices to determine payments, and subjects knew this ahead of time. The experimenter rolled two 10-sided dice at the front of the room to generate a number 1–68.[11] This determined the random question that would be paid. Then, the experimenter rolled the dice again to generate a number 1–100 to resolve any risk in the randomly selected lottery. Subjects were paid the realization from whichever lottery they had chosen in the randomly-selected decision. Therefore, subjects were paid based on exactly one decision they made in the entire experiment.

This payment method, denoted the "random payment selection" (RPS) mechanism, has been used in many binary choice experiments. As discussed in Azrieli et al. (2019), by using the RPS mechanism, we are assuming that *compound independence* holds (Segal, 1990).[12] Brown and Healy (2018) give evidence that compound independence holds when presenting choices on separate screens, as we do in our experiment. Segal (1990) shows that compound independence and reduction of compound lotteries together imply *mixture independence*, which is the form of independence we study. Therefore, by using this payment mechanism, we assume that individuals do not always satisfy reduction of compound lotteries, since we observe violations of mixture independence.

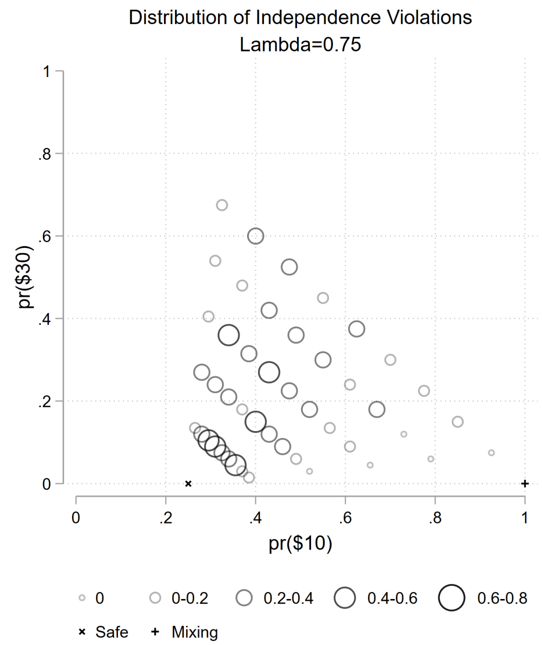
V. Results

We focus our main results on the *certain* comparisons in the Allais Mix treatment. These are questions where subjects chose between a risky lottery and \$20 for sure in the unmixed question, and separately made the same binary comparison when both were mixed with 100% chance of \$10, for three different mixing probabilities $\lambda = \{0.75, 0.50, 0.25\}$. Across all of these binary comparisons, individuals violate independence 25% of the time.[13]

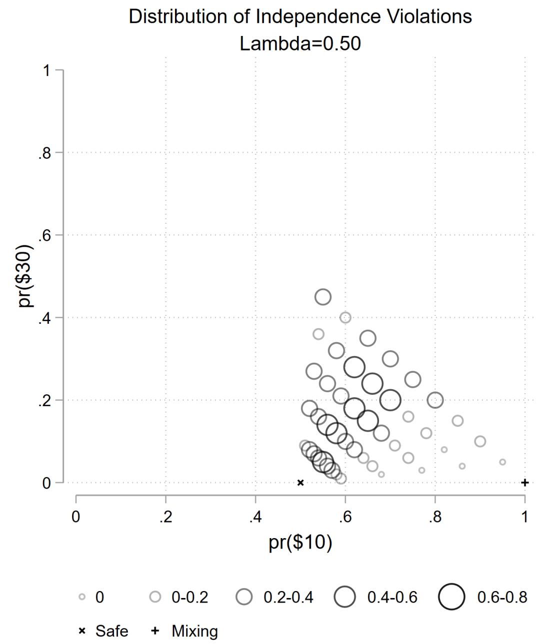
[11] If the number came up larger than 68, she rolled again.

[12] Let A and B be two-stage lotteries over the simple lotteries in our experiment. That is, A=$(\alpha_p, p; \alpha_q, q; ...; \alpha_r, r; ...; \alpha_s, s)$ is a two-stage lottery that gives simple lottery p with probability α_p, lottery q with probability α_q, etc. Let B=$(\alpha_p, p; \alpha_q, q; ...; \alpha_r, t; ...; \alpha_s, s)$, meaning that lottery B differs from lottery A only in that B gives lottery r with probability α_r while A gives lottery t with that same probability. Compound independence says that A is preferred to B if and only if r is preferred to t.

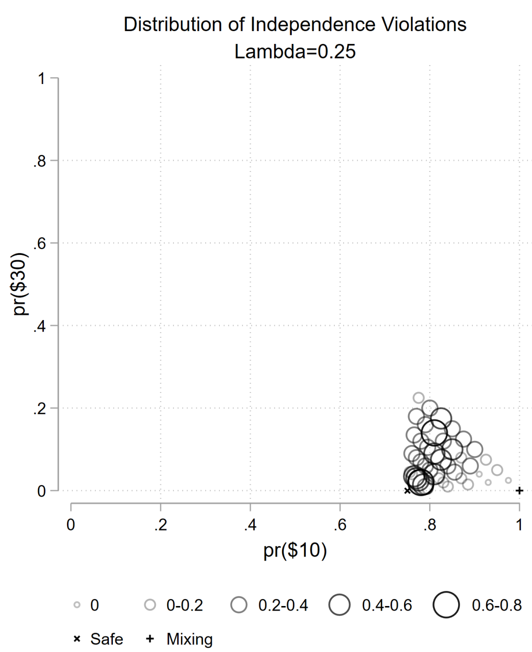
[13] This percentage likely would change as we change payoffs, unmixed lotteries, etc. Therefore, we do not emphasize the raw percentage of violations, and leave it to the reader to decide whether this is



(1) $\lambda = 0.75$



(2) $\lambda = 0.50$



(3) $\lambda = 0.25$

Figure II: Independence Violations in the Simplex

Notes: Figures show percentage of independence violations in the Allais Mix questions, compared with δ_{20}. Size of bubbles denote frequency of violations, with percentages as indicated in the legend.

Figure II shows the violations of independence in the simplex, separated by mixing probability. We find higher violations of independence as λ decreases.[14] We find individuals violate independence in 22% of decisions when $\lambda = 0.75$, 26% when $\lambda = 0.50$, and 27% when $\lambda = 0.25$ (Wilcoxon ranksum p-values, 0.75 vs. 0.25 $p = 0.007$, 0.75 vs. 0.50 $p = 0.074$, 0.50 vs. 0.25 $p = 0.361$). We also see that violations appear more common for risky lotteries with higher expected value (lotteries to the northwest). These are lotteries where individuals are more likely to have chosen the risky option in the unmixed question. We formalize this conjecture below, showing that individuals indeed violate independence more when the risky lottery is preferred to certainty.

Certainty Effect vs. Reverse Certainty Effect

We denote a violation of independence as a "reverse certainty effect" (RCE) violation when individuals prefer the risky lottery to δ_{20} in the unmixed question but reverse their preference in the mixed question. We refer to the opposite as a "certainty effect" (CE) violation, when individuals prefer δ_{20} to the risky lottery in the unmixed question but reverse their preference in the mixed question. Table I presents our main results. In the second column, we compare the aggregate percentages of CE and RCE violations. Contrary to popular belief, we find that RCE violations are twice as common as CE violations (66% vs. 33%, Fisher-Pitman permutation test $p < 0.001$).

	% of Independence Violations	% Violating Independence
Chose Risky in Unmixed (n=419)	66%	39%
Chose Safe in Unmixed (n=578)	33%	15%
p-value	< 0.001	< 0.001

Table I: Percentage of Independence Violations Conditional on Choice in Unmixed Question

Note: Results are separated by choice in the unmixed question. $p \succeq \delta_{20}$ indicates instances where individuals chose risk over certainty, while $p \preceq \delta_{20}$ indicates the opposite.

The last column of Table I looks at the likelihood of violating independence conditional on an individual's choice in the unmixed question. Here, we find the complementary result. Individuals violate independence in 39% of binary comparisons

a large number or not.

[14] This could be because the lotteries converge as λ decreases, so they become closer to one another in expected value. Alternatively, it could be because they are least similar to the original lotteries.

where they chose the risky option in the unmixed question. On the contrary, they violate independence in only 15% of binary comparisons where they chose δ_{20} in the unmixed question. Therefore, we find that individuals are significantly more likely to violate independence in questions where they prefer risk to certainty, rather than the opposite (Fisher-Pitman permutation test, $p < 0.001$). In the Appendix, we show that this result is not driven by having more questions where individuals prefer the safe alternative. Results hold to the same extent when we look at a sub-sample balanced across risky-safe choices.

Our main result is surprising in light of a large literature on the certainty effect. We find that independence violations are much more common when individuals prefer risk to certainty in the absence of mixing. This is exactly the opposite of Allais's intuition, which hypothesized that independence violations would be driven by a "preference for security." Instead, two-thirds of our violations result when a risky lottery is preferred to certainty.

The Role of Certainty

The original intuition of the certainty effect claimed that "certainty" held a fundamentally different appeal from "near certainty." We test this by analyzing differences in independence violations when comparisons involve true certainty (δ_{20}) versus near-certain ($q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$) options. We find the overall percentage of independence violations is slightly but significantly higher under certainty (25% vs. 22%, Fisher-Pitman Permutation Test, $p = .00415$).

	Percentage of Independence Violations	
	Near-Certain	Certain
Chose Risky in Unmixed	36%	39%
Chose Safer in Unmixed	11%	15%
p-value	< 0.001	< 0.001

Table II: Percentage of Independence Violations Conditional on Choice in Unmixed Question, by Near-Certain and Certain

Note: Results are separated by choice in the unmixed question.

Nevertheless, Table II shows that our main result holds equally under certainty as near certainty: Individuals are nearly three times more likely to violate independence in questions where risk is preferred to certainty or near-certainty.

Dependent Variable: Violation of Independence	
Unmixed Risky	0.858***
	(0.0997)
Certain	0.174**
	(0.0787)
Unmixed Risky × Certain	-0.0801
	(0.112)
No. Observations	6,018
No. Clusters	118

Table III: Probit regression predicting violations of Independence

Notes: The dependent variable is a dummy taking the value of 1 for a violation of independence, 0 otherwise. *Unmixed Risky* is a dummy taking the value of 1 if the individual chose the risky lottery in the unmixed question, 0 otherwise. The *Certain* dummy takes the value of 1 when the unmixed question compared against δ_{20}, 0 for questions compared against q^*. We cluster standard errors at the subject level.

Table III confirms these results in a probit regression. The dependent variable is a dummy taking the value of 1 for a violation of independence, 0 otherwise. Independent variables include an *Unmixed Risky* dummy, taking the value of 1 if the individual chose the risky lottery in the unmixed question, 0 otherwise. We include a *Certain* dummy taking the value of 1 when the unmixed question compared against δ_{20}, 0 for questions compared against q^*. We also include the interaction between these two variables. We cluster standard errors at the subject level.

The results from the regression confirm the conclusions above. Violations of independence are significantly more common when individuals choose the riskier option in the unmixed question. Overall violations are slightly more common under certainty. Regardless, RCE violations prevail.

Individual-Level Results

Our main result holds on an individual level, as well. Figure III shows the percentage of independence violations per subject, broken down by CE and RCE types. That is, for each individual, we separate the 17 unmixed questions they answered according to whether they preferred the risky or safer lottery. Then, we compute the average independence violations, for each individual, within these two sets. We see that the modal subject never violates independence in questions where they preferred the safer lottery to the risky lottery. When they preferred the risky lottery, however, the modal subject violates independence one-third of the time.

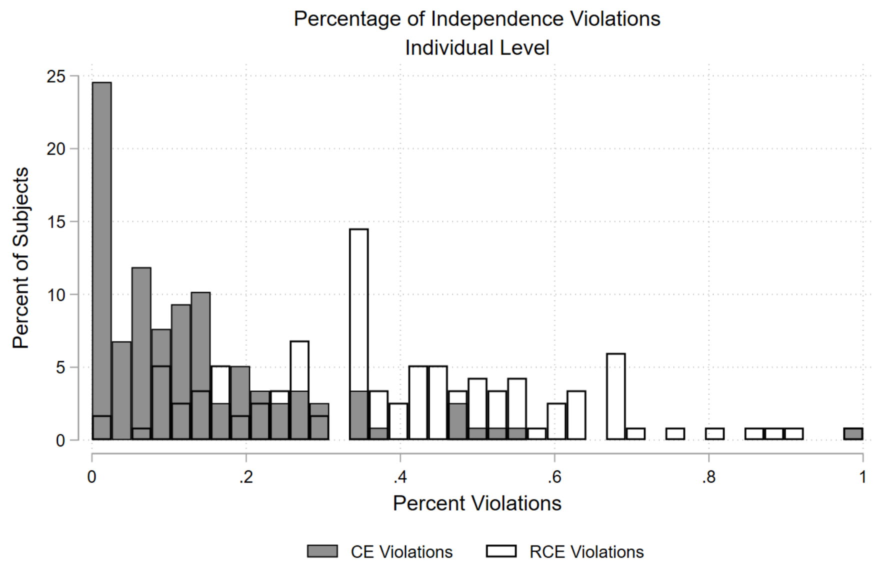


Figure III: Individual-Level Violations by Unmixed Risky vs. Unmixed Safe
Notes: Histogram shows the percentage of questions in which each individual violates the independence axiom, separated out by CE violations (where the individual preferred δ_{20} to the risky lottery) and RCE violations (where they preferred the risky lottery to δ_{20}).

Furthermore, for each subject, we take the difference between their likelihood of RCE and CE violations (RCE - CE). For example, a subject who violated independence in 25% of questions where they preferred the risky lottery over δ_{20} and violated independence in 15% of questions where they preferred δ_{20} over the risky lottery would give a difference of 0.10. Figure IV shows the results. A large majority of subjects (81%) demonstrate a positive difference in violations, meaning that a majority of our subjects express more RCE than CE violations. Therefore, we conclude that RCE violations are more common both in aggregate and on an individual-level.

In the Appendix, we also show that the prevalence of RCE violations is not driven by individual risk preferences. We use a jackknife procedure to calculate the percentage of independence violations for subjects who chose the risky versus safer lottery in each question, repeated for all questions in the simplex. We find no average differences across these sub-samples.

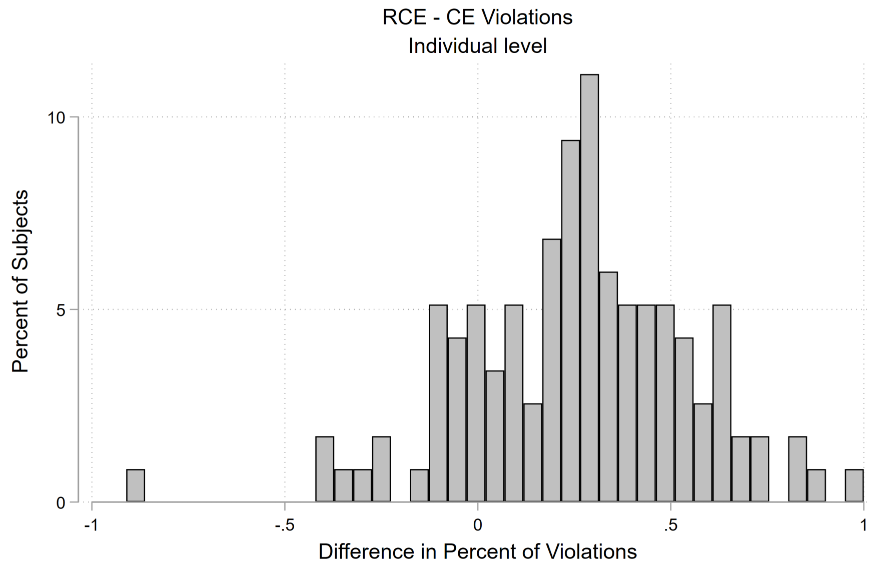


Figure IV: Individual-Level Difference Between Percentage of RCE and CE Violations

Notes: Histogram shows individual-level difference in the percentage of questions in which the individual violates the independence axiom in preferring the risky lottery over δ_{20} minus the percentage of questions in which the individual violates the independence axiom in preferring δ_{20} over the risky lottery.

A More Rigorous Test

Given that each individual answers four questions all linked by independence—$\lambda = \{1, 0.75, 0.50, 0.75\}$—we can analyze violations of independence using all four choices as a unit of observation. We define an individual's four choices by a string of four letters, each R (riskier) or S (safer). The first letter represents their choice in the unmixed question ($\lambda = 1$), the second in the $\lambda = 0.75$ mixed question, the third in the $\lambda = 0.50$ question, and the last in the $\lambda = 0.25$ question. Thus, the string represents choices as we move towards the bottom right of the simplex.

Independence requires the individual choose either R or S in all four questions. This is a more stringent requirement than in our main analysis, as it requires individuals to be consistent in all three binary comparisons. Nevertheless, these are the two most common patterns we see, first SSSS followed by RRRR. About 60% of our data falls into one of those two patterns.[15] Consistent with the analysis above,

[15] Recall that just looking at binary comparisons, about 75% of choices were consistent with independence.

Pattern	Near-Certain	Certain
SSSS	449	425
RRRR	191	159
RSSS	71	72
RRRS	41	36
RRSS	38	45
RRSR	32	38
SSSR	29	43
RSRR	24	24
RSRS	22	19
SSRS	22	18
RSSR	21	26
SRRR	18	29
SRSS	18	19
SSRR	15	12
SRSR	12	17
SRRS	6	15

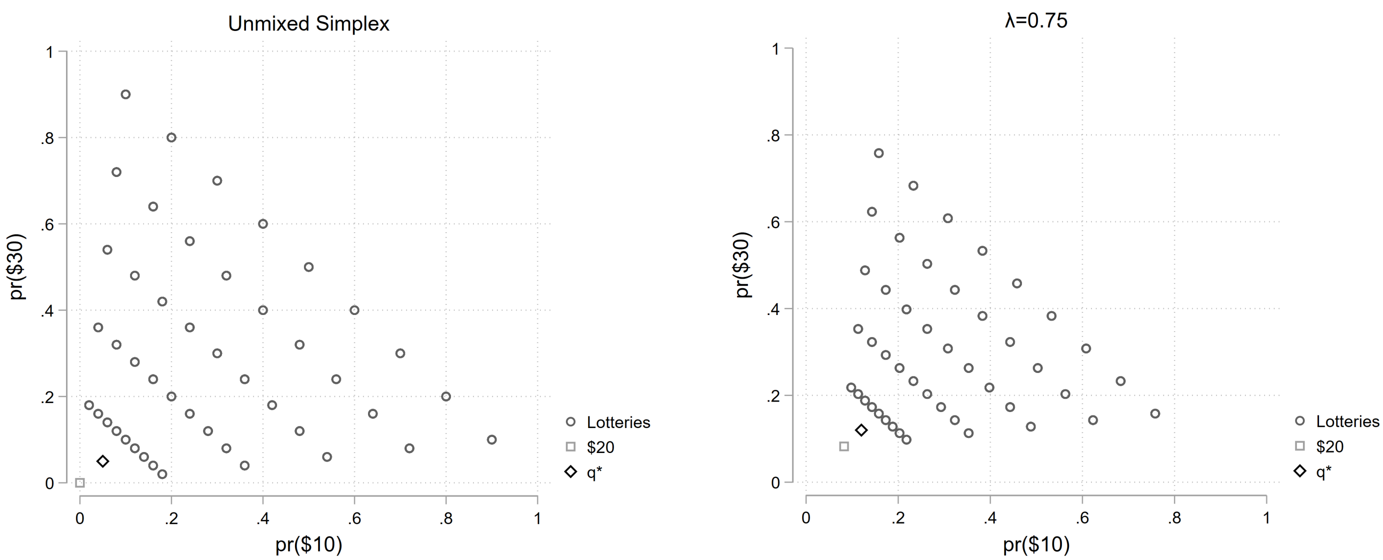
Table IV: Pattern of Choices Per Question

however, the next most common patterns involve choosing the riskier option in the unmixed question and choosing the safer alternative as λ decreases. In the most common independence violation, RSSS, individuals choose the riskier option only in the unmixed question but then choose the safer option in all mixtures. In the second most common violation, RRSS, individuals choose the riskier option in the unmixed question and $\lambda = 0.75$, but choose the safer option in the $\lambda = 0.50, 0.25$ mixtures. All patterns can be found in Table IV. We find no significant difference across certain and uncertain questions (Chi-square $p = 0.351$).

Analyzing the data from all four questions also allows us to test whether violations of independence are less common in the interior of the simplex. Allais hypothesized that "'far from certainty,' individuals act as expected utility maximizers" (Allais, 1953, translated by Andreoni and Sprenger, 2010). If this were the case, we would only see independence violations of the SRRR and RSSS types, where individuals act consistently with independence in the $\lambda = 0.75, 0.50, 0.25$ mixtures but might violate it near certainty. Instead, we find that these patterns make up only 24% of independence violations in the certain comparisons in our sample. Over three quarters of choice patterns violating independence involve violations in questions that lie strictly on the interior of the simplex.[16]

Furthermore, the SRRR pattern most closely associated with the certainty effect is very rare. These are instances in which individuals choose the safe option when it is certain or near-certain, but then reverse their preferences when both alternatives move away from certainty. Conditional on choosing S in the $\lambda = 1$ question, only 17% of violations follow this pattern.

Robustness to the Mixing Lottery



[16]All near-certain violations are also strictly interior.

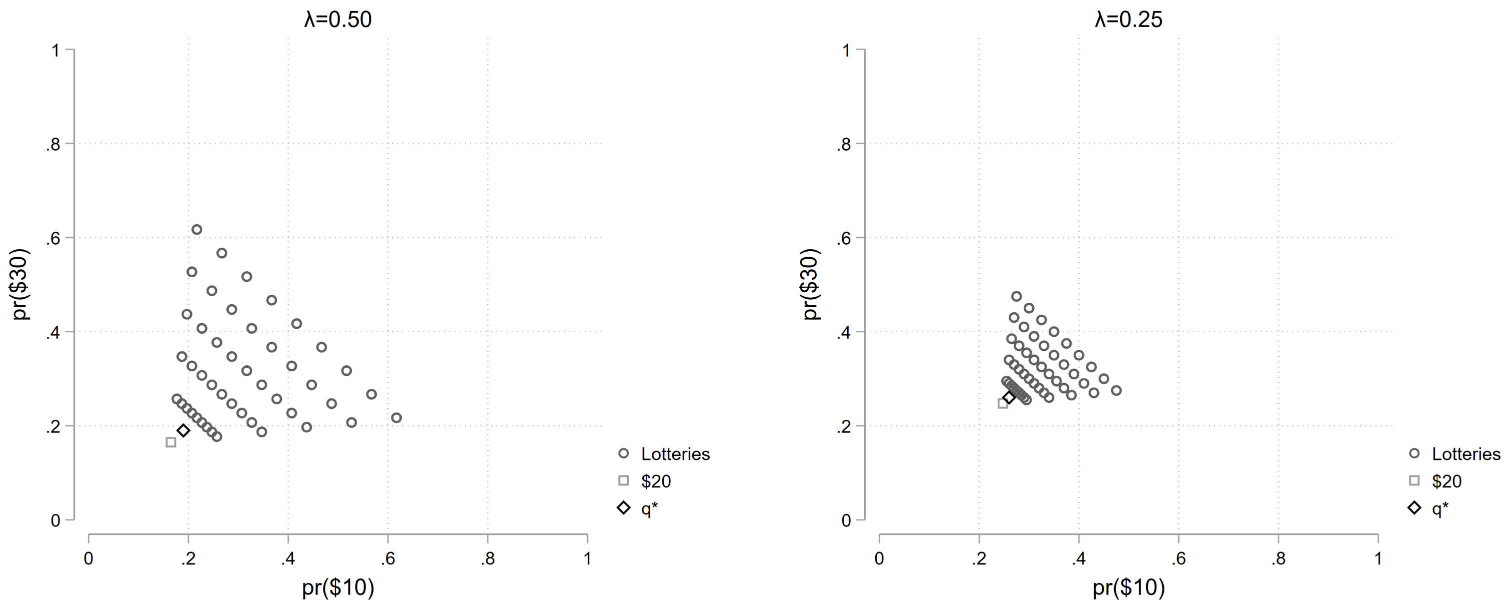


Figure V: Questions in the Middle Mixture Treatment

We test the robustness of our results to the choice of r, the mixing lottery in the definition of the independence axiom: $p \succeq q \Leftrightarrow \lambda p + (1-\lambda)r \succeq \lambda q + (1-\lambda)r$. In our main treatment, we chose r to be ($10, 1), or 100% chance of the lowest payoff. This is closest in spirit to the original Allais paradox, where the lotteries were mixed with a large chance of receiving $0. In our robustness sessions, we conducted exactly the same experiment, except we mixed all lotteries instead with the midpoint of the simplex, ($\$30, \frac{1}{3}; \$20, \frac{1}{3}; \$10, \frac{1}{3}$). We chose this point so that mixing would converge to a different area of the simplex, one which has not been studied in detail based on our review of the literature. Other than the choice of r, all procedures in these sessions followed identically to those in the main session.

Overall, individuals violate independence in 23% of questions, which is slightly but significantly higher than in our original treatment (23% vs. 20%, Chi-square $p < 0.001$). Figures XI and XII in the Appendix show the distribution of violations across the simplex.

Under certainty, again we see RCE violations are more common than CE violations. Figure VI shows that two thirds of violations under certainty are driven by individuals who choose the riskier option in the unmixed question ($p < 0.001$). On the other hand, violations near certainty are driven equally by those choosing the riskier and safer option in the unmixed question (permutation test $p = 0.146$). While our main treatment did not show any significant differences between certainty and near-certainty, our robustness treatment suggests that certainty does have an impact outside the Allais paradigm.

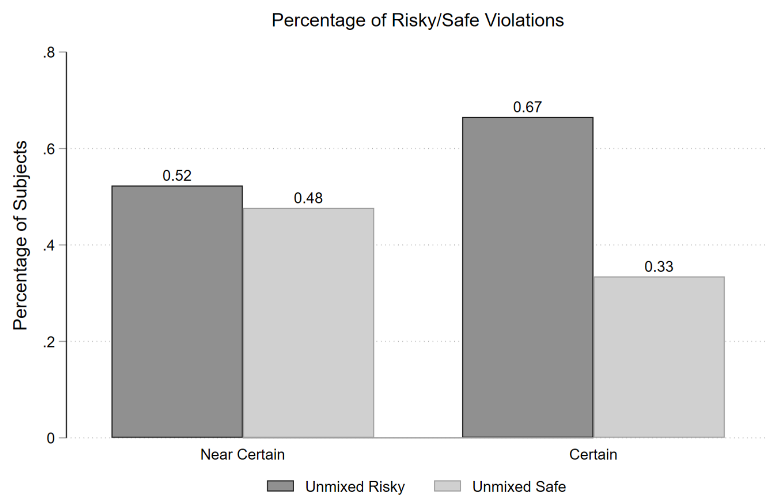


Figure VI: Percentage of RCE/Safe Choices Among Violations
Notes: "Unmixed Risky" refers to questions in which the individual chose the risky lottery over the safer lottery in the unmixed question. "Unmixed Safe" refers to the opposite. "Near Certain" are comparisons with $q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$ while "Certain" are comparisons with δ_{20}.

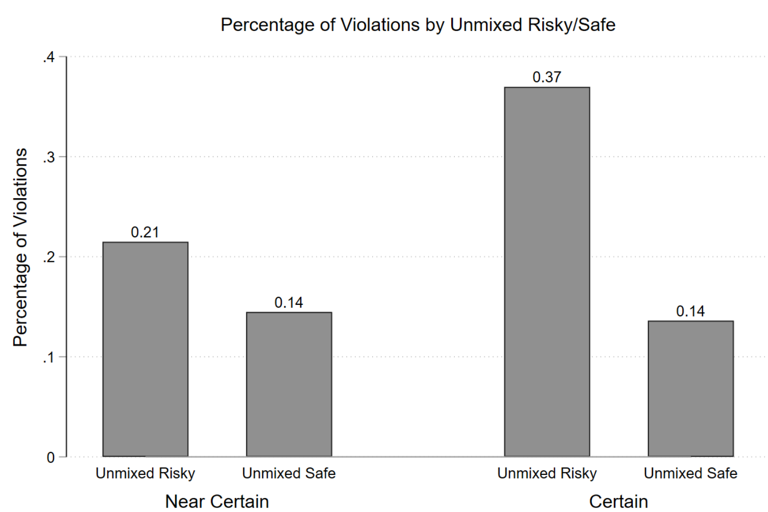


Figure VII: Percentage of Violations by RCE/Safe
Notes: "Unmixed Risky" refers to questions in which the individual chose the risky lottery over the safer lottery in the unmixed question. "Unmixed Safe" refers to the opposite. "Near Certain" are comparisons with $q^* = (\$30, 0.05; \$20, 0.90; \$10, 0.05)$ while "Certain" are comparisons with δ_{20}.

Figure VII confirms this further. In the case of certainty, individuals who choose the risky option in the unmixed question are much more likely to violate independence than those who choose the safe option. Near certainty, however, this difference is much smaller ($p < 0.001$ for both). Therefore, it seems that the choice of the mixing lottery contributes to the effect of certainty.

Its important to note, however, that while certainty does play a role, it is in the direction exactly *opposite* the consensus in the literature. The literature would suggest that certainty plays a role in inducing violations where certainty is preferred to risk in the absence of mixing. Our results suggest the opposite effect is much stronger: Violations of independence are most common when risk is preferred to certainty.

VI. Discussion

We study the independence axiom and how violations of independence interact with a preference for certainty. Contrary to a prominent thread in the literature, we find that violations of independence in our data are not predominantly driven by this preference for certainty. Instead, violations are more common when individuals prefer risk to certainty. We find this is also true when we move slightly away from certainty.

Our results are surprising in light of the large literature following up on the original Allais Paradox counter-examples to independence. The certainty effect is well-documented and is one of the primary pieces of evidence motivating new theoretical models. Our paper aims to provide a more structured analysis to document violations of the independence axiom near certainty. Our results suggest caution in attributing violations of independence to the certainty effect primarily, but more evidence is required before making general statements on where and when to expect violations of expected utility.

We document consistent patterns of behavior, but leave open the question of what drives these preferences. In particular, correlation of independence violations with measures such as IQ, cognitive reflection test (CRT) scores, etc. remain an interesting open question for future research. We also leave open the questions of how these patterns of violations change with payment amounts and other parameters of the decision environment.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine,. *Econometrica*, 21(4):503–546.

Andreoni, J. and Harbaugh, W. T. (2010). Unexpected utility: Experimental tests of five key questions about preferences over risk. *Working Paper*.

Andreoni, J. and Sprenger, C. (2010). Certain and uncertain utility: The allais paradox and five decision theory phenomena. *Working Paper*.

Andreoni, J. and Sprenger, C. (2011). Uncertainty equivalents: Testing the limits of the independence axiom. *NBER Working Paper*.

Andreoni, J. and Sprenger, C. (2012). Risk preferences are not time preferences. *American Economic Review*, 102(7):3357–3376.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2019). Incentives in experiments with objective lotteries. *Experimental Economics*, pages 1–29.

Blavatskyy, P., Ortmann, A., and Panchenko, V. (2015). Now you see it, now you don't: How to make the allais paradox appear, disappear, or reverse. *UNSW Business School Working Paper*.

Blavatskyy, P. R. (2010). Reverse common ratio effect. *Journal of Risk and Uncertainty*, 40(3):219–241.

Blavatskyy, P. R. (2013). The reverse allais paradox. *Economics Letters*, 119(1):60–64.

Brown, A. L. and Healy, P. J. (2018). Separated decisions. *European Economic Review*, 101:20–34.

Camerer, C. (1995). Individual decision making. *Handbook of experimental economics*.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2015). Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728.

Cerreia-Vioglio, S., Dillenberger, D., and Ortoleva, P. (2020). An explicit representation for disappointment aversion and other betweenness preferences. *Forthcoming, Theoretical Economics*.

Conlisk, J. (1989). Three variants on the allais example. *The American Economic Review*, pages 392–407.

Dekel, E. (1986). An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic theory*, 40(2):304–318.

Dillenberger, D. (2010). Preferences for one-shot resolution of uncertainty and allais-type behavior. *Econometrica*, 78(6):1973–2004.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2).

Greiner, B. (2004). An online recruitment system for economics experiments.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59(3).

Halevy, Y. (2008). Strotz meets allais: Diminishing impatience and the certainty effect. *American Economic Review*, 98(3):1145–1162.

Hong, C. S. and Waller, W. S. (1986). Empirical tests of weighted utility theory. *Journal of mathematical psychology*, 30(1):55–72.

Humphrey, S. and Verschoor, A. (2004). The probability weighting function: Experimental evidence from uganda, india, and ethiopia. *Economics Letters*, 84:419–425.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Schmidt, U. (1998). A measurement of the certainty effect. *Journal of Mathematical Psychology*, 42:32–47.

Segal, U. (1990). Two-stage lotteries without the reduction axiom. *Econometrica*, 58:349–377.

Starmer, C. (1992). Testing new theories of choice under uncertainty using the common consequence effect. *The Review of Economic Studies*, 59(4):813–830.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, 38(2):332–382.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.

A. Additional Results

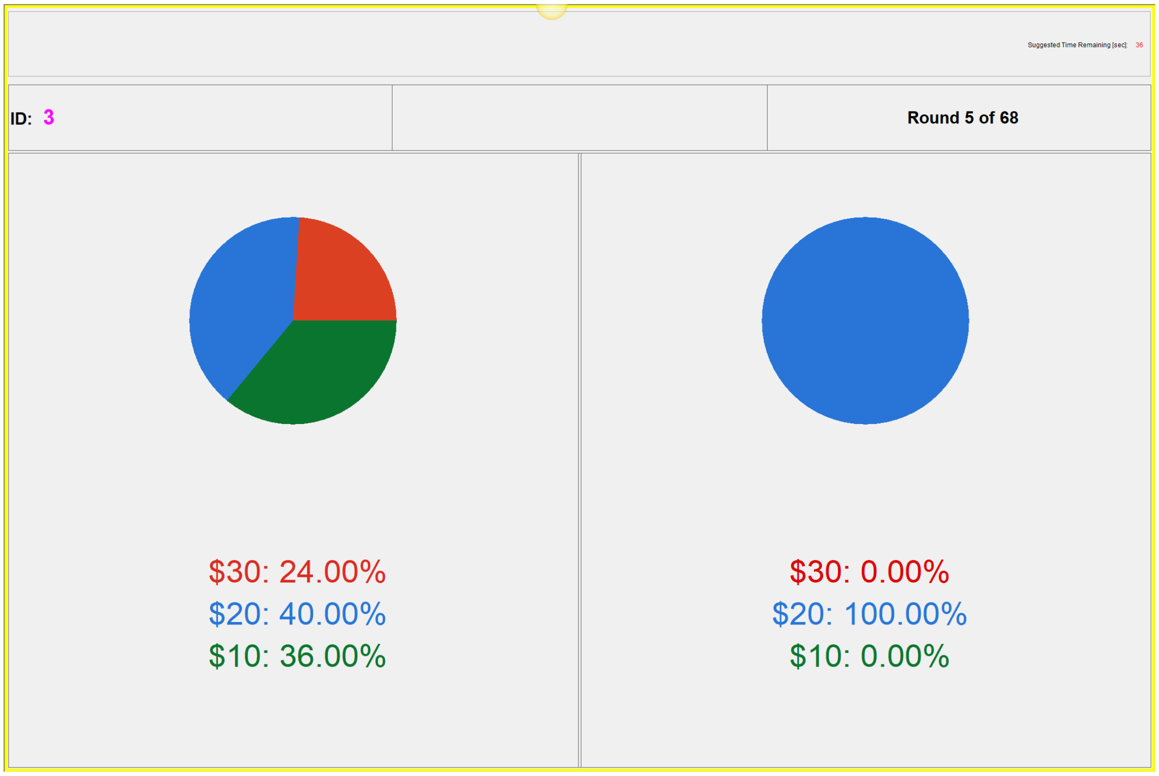


Figure VIII: Screenshot of subjects' display during the experiment

Risk Preferences

In this section, we look to see whether our main results are driven by individual risk preferences. Our main result could emerge from one of two behavioral phenomenon. It could be the case that *any* individual, when choosing a risky lottery over a sure amount, is more likely to violate independence than when choosing the sure amount over the risky lottery. On the other hand, it could be that more risk-seeking individuals, who are more likely to choose the risky lottery over the sure amount, are more likely to violate independence. We rule out the latter explanation using the following non-parametric procedure. For each of the original 45 questions of p vs. \$20, we split the sample into those who chose the risky lottery and those who chose the sure amount. Then, we look at average independence violations *in all other questions* for these two types of individuals. If independence violations are driven by risk preferences, we should see more violations for risky types than for safe types. Figure IX shows the histogram of violations. A Wilcoxon sign-rank test shows no significant differences in these distributions ($p = 0.569$). Therefore, it is not the case that our pattern of violations is driven by individual heterogeneity in risk preferences. Instead, it seems to be the case that individuals will be more likely to violate the independence axiom in any question where they prefer a risky lottery over a sure amount.

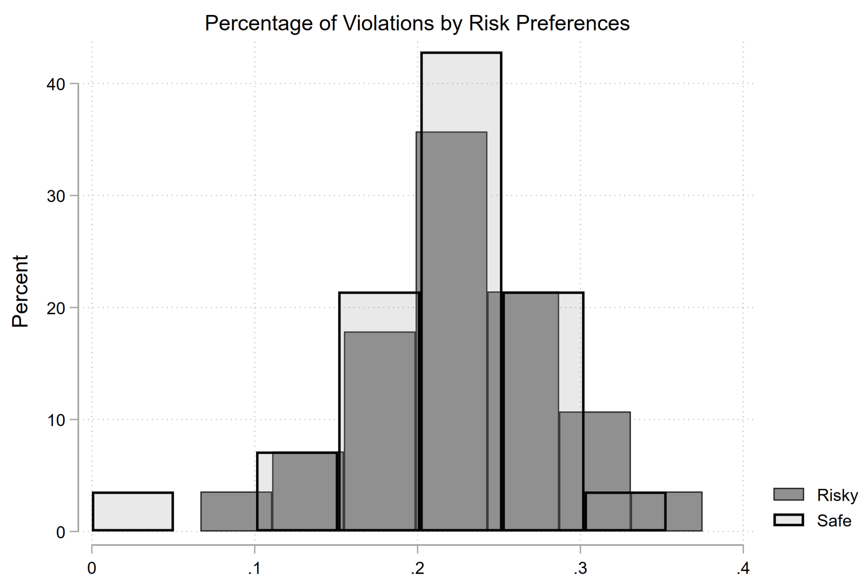


Figure IX: Histogram of Independence Violations by Risk Preferences

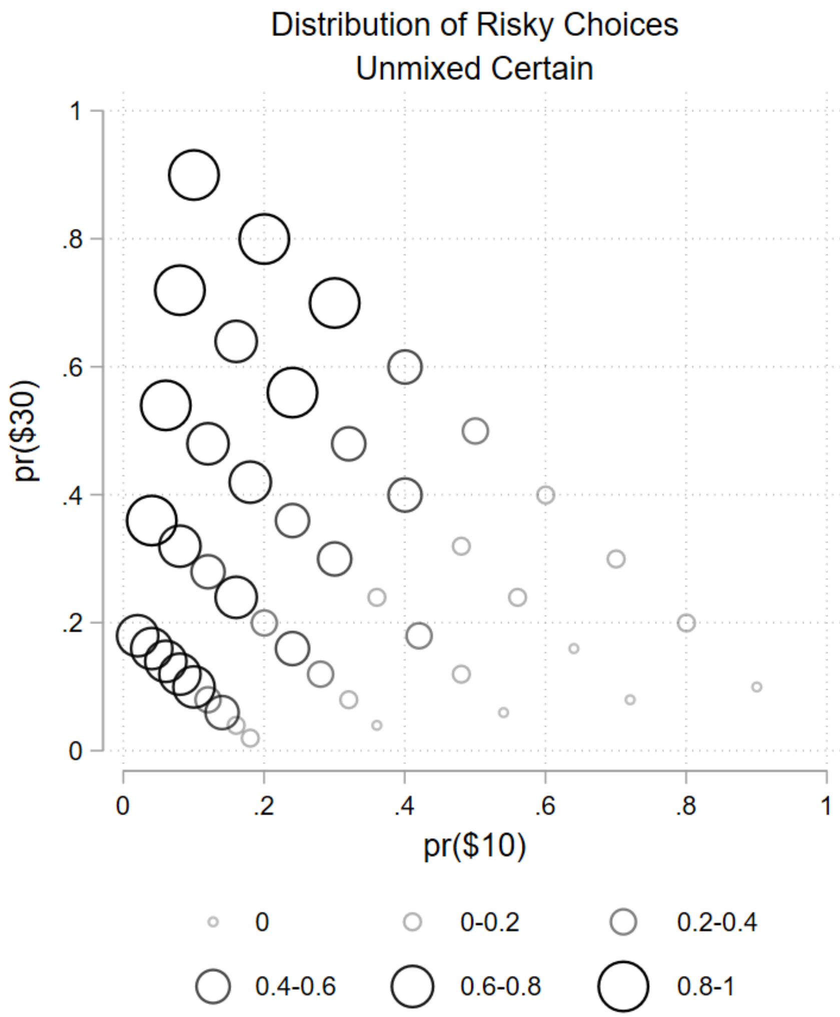
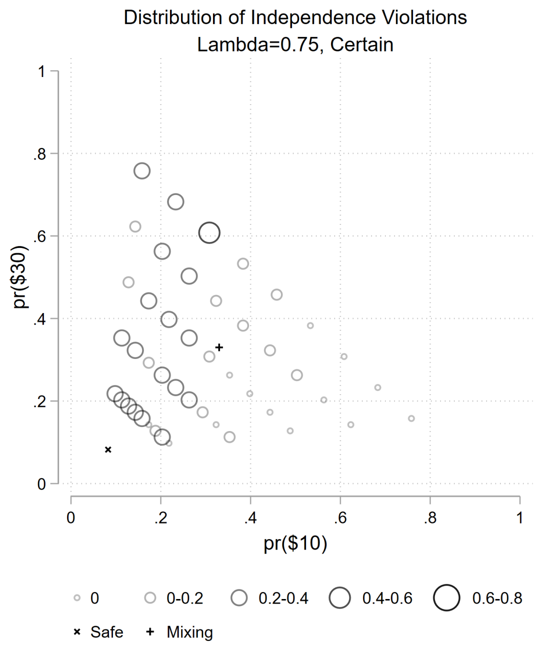
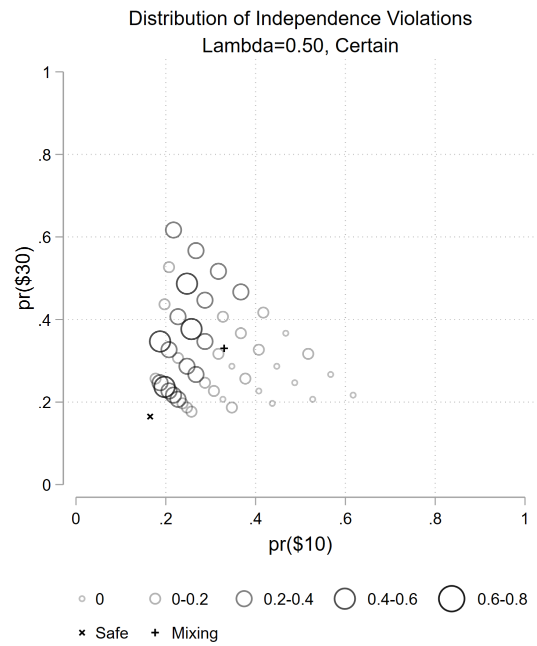


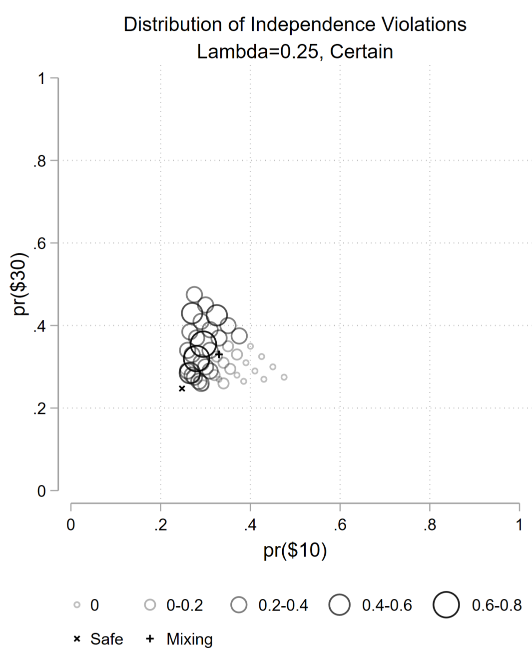
Figure X: Distribution of Risky Choices in the Unmixed Questions



(1) $\lambda = 0.75$



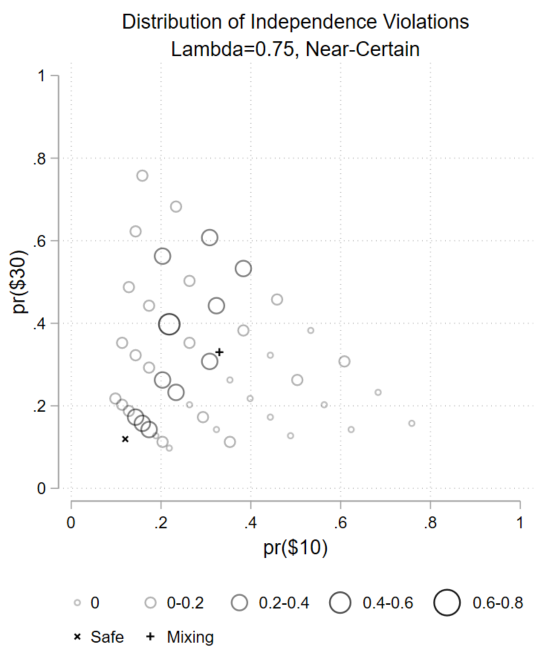
(2) $\lambda = 0.50$



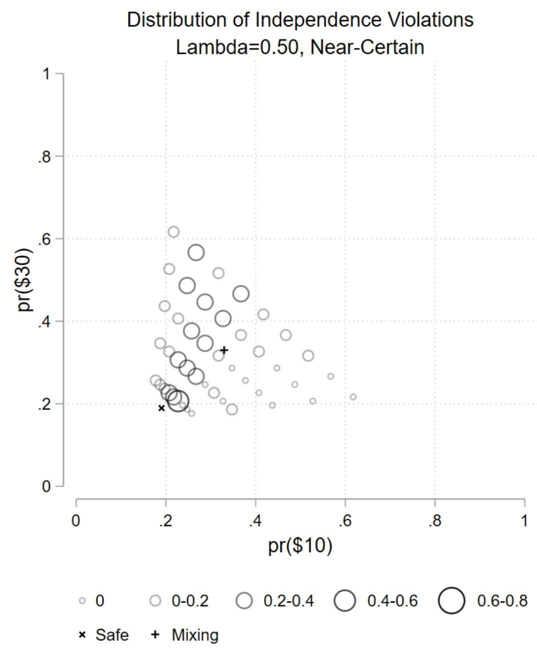
(3) $\lambda = 0.25$

Figure XI: Independence Violations in the Simplex

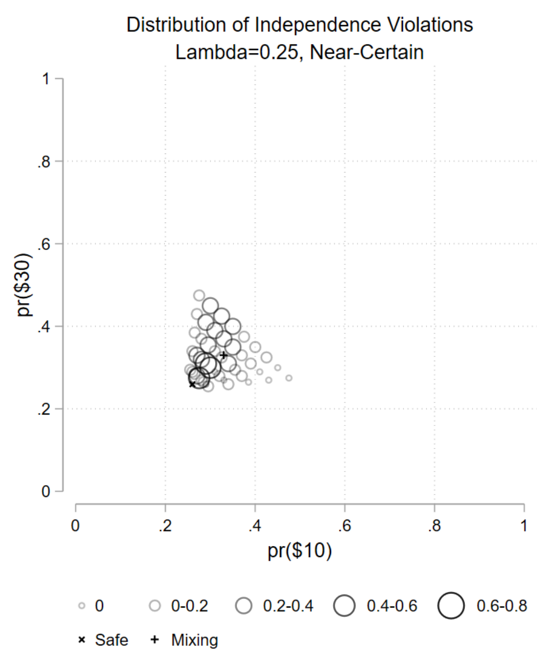
Notes: Figures show percentage of independence violations in the Middle mix questions, compared with certainty. Size of bubbles denote frequency of violations.



(1) $\lambda = 0.75$



(2) $\lambda = 0.50$



(3) $\lambda = 0.25$

Figure XII: Independence Violations in the Simplex

Notes: Figures show percentage of independence violations in the Middle mix questions, compared with q∗. Size of bubbles denote frequency of violations.

Sample Balance

One might worry that these results are driven "mechanically" by risk preferences. If there are more points where individuals prefer the safer option in the unmixed question, then it might be less likely for violations to come from these points simply as a statistical artifact. We chose our simplex to be "balanced" in a natural way, but we need to ensure it is "behaviorally balanced" in order to make the claim that one type of violation is more common than another. First, we note that individuals chose the safer alternative in 57% of the unmixed questions. This is significantly larger than the 43% of risky choices (Fisher-Pitman $p < 0.001$). This could bias our results, since it suggests that there are more questions where the safe thing is "obviously" preferred than there are questions where the risky thing is "obviously" preferred. These might be questions where we would expect the fewest violations of independence. We correct for this by identifying these lotteries and dropping them from the analysis.

In particular, recall that the lotteries in our design lie on nine different rays originating from the origin in Figure I. We calculate the percentage of subjects (92%) choosing the riskier alternative for unmixed questions lying on the steepest ray. Then, we drop rays where the percentage of subjects choosing the *safer* alternative for unmixed questions lies at or above 92%.[17] In this remaining subset of questions, there should be no more questions where the safe thing is "obviously" preferred than there are questions where the risky thing is "obviously" preferred. In other words, we ensure the simplex is "behaviorally balanced" to account for subjects' risk preferences. On this subset, 55% of subjects choose the riskier option in the unmixed question, now slightly unbalanced in the *opposite* direction. Therefore, if we find that RCE violations still prevail, we can be confident that the result is not driven by an unbiased sample.

On this restricted sample, we see the same results hold. Overall, we find that 73% of violations near certainty and 66% of violations with certainty come from individuals who chose the risky option in the unmixed question. As Figure XIII demonstrates, individuals are significantly more likely to violate independence in situations where they chose risk over certainty (near certainty).

[17] In doing so, we drop the two flattest rays.

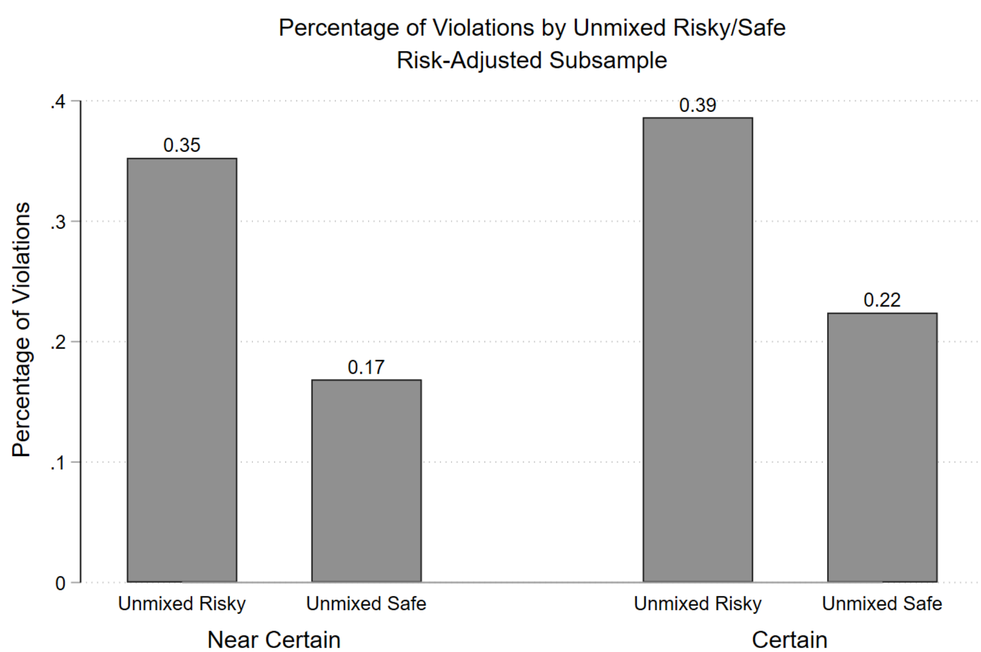


Figure XIII: Percentage of Violations by Choice in the Unmixed Question on our Behaviorally Balanced Sample